270100 - ADEI - Data Analysis and Information Exploitation

Coordinating unit: 270 - FIB - Barcelona School of Informatics
Teaching unit: 715 - EIO - Department of Statistics and Operations Research
Academic year: 2016
Degree: BACHELOR'S DEGREE IN INFORMICS ENGINEERING (Syllabus 2010). (Teaching unit Optional)
ECTS credits: 6
Teaching languages: Catalan, Spanish

Teaching staff

Coordinator: - Lidia Montero Mercadé (lidia.montero@upc.edu)
Others: - Mari Paz Linares Herreros (mari.paz.linares@upc.edu)

Prior skills

Students must have completed a course in probability and statistics and a course on business and economic environment

Requirements

- Prerequisite EEE
- Prerequisite BD
- Prerequisite PE

Degree competences to which the subject contributes

Specific:

CSI2.1. To demonstrate comprehension and apply the management principles and techniques about quality and technological innovation in the organizations.
CSI2.3. To demonstrate knowledge and application capacity of extraction and knowledge management systems.

General:

G3. THIRD LANGUAGE: to know the English language in a correct oral and written level, and accordingly to the needs of the graduates in Informatics Engineering. Capacity to work in a multidisciplinary group and in a multi-language environment and to communicate, orally and in a written way, knowledge, procedures, results and ideas related to the technical informatics engineer profession.
G9. PROPER THINKING HABITS: capacity of critical, logical and mathematical reasoning. Capacity to solve problems in her study area. Abstraction capacity: capacity to create and use models that reflect real situations. Capacity to design and perform simple experiments and analyse and interpret its results. Analysis, synthesis and evaluation capacity.

Teaching methodology

Learning the course consists of three distinct phases:
1. Acquisition of specific knowledge through the study of literature and material provided by teachers.
2. The acquisition of skills in specific techniques of data analysis and exploitation of information and
3. Integration of knowledge, skills and competencies (specific and generic) by solving a real Case Study.

In theory classes serve to expose the foundations of methodologies and techniques of the subject. The laboratory classes are used to learn the use of specific techniques for solving problems, using appropriate informatics tools, in this sense, students first must repeat the problem solved previously by the teachers and then solve a similar one. While the case study, are settled in groups in selflearning hours, and serves to put into practice the knowledge, skills and competences in solving a real case of ADEI.
Learning objectives of the subject

1. Learn how to identify the three levels of decision making in a company
2. Learn how to make a report on data quality
3. Control Quality
4. Continuous process control
5. Control of discrete indicators
6. Determining the drivers of continuous response
7. Diagnosis of a statistical model
8. Modelling of discrete choices
9. Modelling the propensity
10. Analysis of databases. Determination of the significant characteristics of groups of individuals.
11. Concept and measurement of intangibles in a company
12. Multivariate information visualization
13. Clustering
15. Statistical tools for support decision making

Study load

<table>
<thead>
<tr>
<th>Total learning time: 150h</th>
<th>Theory classes: 30h</th>
<th>20.00%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Practical classes: 0h</td>
<td>0.00%</td>
</tr>
<tr>
<td></td>
<td>Laboratory classes: 30h</td>
<td>20.00%</td>
</tr>
<tr>
<td></td>
<td>Guided activities: 6h</td>
<td>4.00%</td>
</tr>
<tr>
<td></td>
<td>Self study: 84h</td>
<td>56.00%</td>
</tr>
</tbody>
</table>
### Content

<table>
<thead>
<tr>
<th>Block</th>
<th>Description</th>
<th>Degree competences to which the content contributes:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bloc1</td>
<td>Levels of corporate decision</td>
<td></td>
</tr>
<tr>
<td>Block 2</td>
<td>Summary description and data quality</td>
<td></td>
</tr>
<tr>
<td>Block 3</td>
<td>Statistical Modeling</td>
<td></td>
</tr>
<tr>
<td>Block 4</td>
<td>Multivariate Data Analysis and intangible measurement</td>
<td></td>
</tr>
<tr>
<td>Block 5</td>
<td>Clustering and profiling</td>
<td></td>
</tr>
</tbody>
</table>
### Planning of activities

<table>
<thead>
<tr>
<th>Activity</th>
<th>Hours</th>
<th>Specific objectives</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Handing in of practical work 1</strong></td>
<td>9h</td>
<td>1, 2, 3, 6, 7, 8</td>
</tr>
<tr>
<td><strong>Presentation of the Case of Study</strong></td>
<td>9h</td>
<td>1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15</td>
</tr>
<tr>
<td><strong>Handing in of practical work 2</strong></td>
<td>9h</td>
<td>10, 11, 12, 13, 14</td>
</tr>
<tr>
<td><strong>Quiz Blocks 4 and 5</strong></td>
<td>9h</td>
<td>1, 10, 11, 12, 13, 14</td>
</tr>
<tr>
<td><strong>Quiz blocks 2 and 3</strong></td>
<td>9h</td>
<td>1, 2, 3, 6, 7, 8</td>
</tr>
<tr>
<td><strong>Block 1. Levels of corporate decision</strong></td>
<td>5h</td>
<td>Theory classes: 2h</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Practical classes: 0h</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Laboratory classes: 2h</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Guided activities: 0h</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Self study: 1h</td>
</tr>
</tbody>
</table>
### Description:
It presented the three levels of decision making in companies. What are the main business processes and how is stored the generated data.

### Specific objectives:
1

### Block 2. Description and quality control data

<table>
<thead>
<tr>
<th>Hours</th>
<th>7h</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theory classes</td>
<td>2h</td>
</tr>
<tr>
<td>Practical classes</td>
<td>0h</td>
</tr>
<tr>
<td>Laboratory classes</td>
<td>1h</td>
</tr>
<tr>
<td>Guided activities</td>
<td>1h</td>
</tr>
<tr>
<td>Self study</td>
<td>3h</td>
</tr>
</tbody>
</table>

**Description:**
Problems in data quality: This is seen in the Case Study or problems that may present data: inconsistency, redundancy, Missing data, Outliers. How do I report data quality. What is the standardization of data.

**Specific objectives:**
2

### Block 2. Treatment of random variability

<table>
<thead>
<tr>
<th>Hours</th>
<th>7h</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theory classes</td>
<td>2h</td>
</tr>
<tr>
<td>Practical classes</td>
<td>0h</td>
</tr>
<tr>
<td>Laboratory classes</td>
<td>2h</td>
</tr>
<tr>
<td>Guided activities</td>
<td>0h</td>
</tr>
<tr>
<td>Self study</td>
<td>3h</td>
</tr>
</tbody>
</table>

**Description:**
Principles of continuous improvement in quality. Definition of indicators and statistical variability. Methodology Operational Control: historical variability

**Specific objectives:**
2, 3, 15

### Block 2. Data visualisation

<table>
<thead>
<tr>
<th>Hours</th>
<th>7h</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theory classes</td>
<td>2h</td>
</tr>
<tr>
<td>Practical classes</td>
<td>0h</td>
</tr>
<tr>
<td>Laboratory classes</td>
<td>2h</td>
</tr>
<tr>
<td>Guided activities</td>
<td>0h</td>
</tr>
<tr>
<td>Self study</td>
<td>3h</td>
</tr>
</tbody>
</table>

**Description:**
Type of Data Collection and applicability to operational control. Indicators common in continuous process control

**Specific objectives:**
2, 3, 15
## Block 3. Statistical Modeling

<table>
<thead>
<tr>
<th>Hours</th>
<th>Theory classes: 2h</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Practical classes: 0h</td>
</tr>
<tr>
<td></td>
<td>Laboratory classes: 1h</td>
</tr>
<tr>
<td></td>
<td>Guided activities: 1h</td>
</tr>
<tr>
<td></td>
<td>Self study: 3h</td>
</tr>
</tbody>
</table>

### Specific objectives:
6, 15

## Block 3. Training the model

<table>
<thead>
<tr>
<th>Hours</th>
<th>Theory classes: 2h</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Practical classes: 0h</td>
</tr>
<tr>
<td></td>
<td>Laboratory classes: 2h</td>
</tr>
<tr>
<td></td>
<td>Guided activities: 0h</td>
</tr>
<tr>
<td></td>
<td>Self study: 3h</td>
</tr>
</tbody>
</table>

### Specific objectives:
6

## Block 3. Validation of statistical modeling

<table>
<thead>
<tr>
<th>Hours</th>
<th>Theory classes: 2h</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Practical classes: 0h</td>
</tr>
<tr>
<td></td>
<td>Laboratory classes: 2h</td>
</tr>
<tr>
<td></td>
<td>Guided activities: 0h</td>
</tr>
<tr>
<td></td>
<td>Self study: 3h</td>
</tr>
</tbody>
</table>

### Description:
Elements involved in the validation of regression modeling. Values influential and / or outliers

### Specific objectives:
7

## Bloc 3. Statistical Modeling of binary variables

<table>
<thead>
<tr>
<th>Hours</th>
<th>Theory classes: 2h</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Practical classes: 0h</td>
</tr>
<tr>
<td></td>
<td>Laboratory classes: 2h</td>
</tr>
<tr>
<td></td>
<td>Guided activities: 0h</td>
</tr>
<tr>
<td></td>
<td>Self study: 3h</td>
</tr>
</tbody>
</table>

### Specific objectives:
7, 8
### Block 4. Multivariate Data Analysis

**Hours:** 7h  
Theory classes: 2h  
Practical classes: 0h  
Laboratory classes: 1h  
Guided activities: 1h  
Self study: 3h

**Specific objectives:**
11, 12

### Block 4. Principal Component Analysis

**Hours:** 7h  
Theory classes: 2h  
Practical classes: 0h  
Laboratory classes: 2h  
Guided activities: 0h  
Self study: 3h

**Specific objectives:**
11, 12

### Block 4. Measurement of intangibles

**Hours:** 7h  
Theory classes: 2h  
Practical classes: 0h  
Laboratory classes: 2h  
Guided activities: 0h  
Self study: 3h

**Specific objectives:**
11, 12

### Block 4. Practice of Principal Component Analysis

**Hours:** 7h  
Theory classes: 2h  
Practical classes: 0h  
Laboratory classes: 2h  
Guided activities: 0h  
Self study: 3h

**Description:**
Practice Principal Component Analysis, interpretation of the representations obtained. Positioning of the supplementary information.

**Specific objectives:**
11, 12, 15
### Block 5. Clustering

**Hours:** 7h  
Theory classes: 2h  
Practical classes: 0h  
Laboratory classes: 2h  
Guided activities: 0h  
Self study: 3h

**Specific objectives:**
13

**Description:**  
Presentation of the k-means and hierarchical methods.

**Specific objectives:**
13, 15

---

### Block 4. Practice of Clustering

**Hours:** 7h  
Theory classes: 2h  
Practical classes: 0h  
Laboratory classes: 1h  
Guided activities: 1h  
Self study: 3h

**Description:**  
Presentation of the k-means and hierarchical methods.

**Specific objectives:**
13, 15

---

### Block 5. profiling

**Hours:** 9h  
Theory classes: 2h  
Practical classes: 0h  
Laboratory classes: 2h  
Guided activities: 1h  
Self study: 4h

**Specific objectives:**
10, 15
The evaluation of the course integrates the three phases of learning process: knowledge, skills and competencies. The knowledge is assessed by two quizzes, in the middle and last week of the course. If you fail this exam, students may have a final resit. (score T).

The skills assessed from the delivery from 2 to 5 practices relating to the course case study. Each of the blocks 2 and 3 involve a practice that students will perform either individually or in groups of 2, the same for blocks 4 and 5. The average of the scores comes out the L score.

The case study as a whole exercise will be evaluated based on the oral presentation (score P).

In the presentation of case study that generic skills will be assessed. In any case, the presentation of the case study is compulsory.

The final grade will obtained weighing the three scores: Final Mark = 0.4P + 0.3L + 0.3T.

Generic skills will be assessed on the scale: Fail, Pass, Good and Very good (D,C,B and A).

To assess the competence on English, it will be required to have written in English the report on the Case Study, moreover at the beginning of the presentation, the student must do an outline of the work in English as well. Regarding the reasoning competence, it will be assessed from the answers given to the presentation of the Case Study.

### Bibliography

#### Basic: